

# Local convergence analysis of Gauss-Newton's method under majorant condition

O. P. Ferreira\*      M. L. N. Gonçalves<sup>†</sup>      P. R. Oliveira<sup>‡</sup>

March 29, 2010

## Abstract

The Gauss-Newton's method for solving nonlinear least squares problems is studied in this paper. Under the hypothesis that the derivative of the function associated with the least square problem satisfies a majorant condition, a local convergence analysis is presented. This analysis allow us to obtain the optimal convergence radius, the biggest range for the uniqueness of solution, and to unify two previous and unrelated results.

**Keywords:** Nonlinear least squares problems; Gauss-Newton's method; Majorant condition; Local convergence.

## 1 Introduction

The Gauss-Newton's method is one of the most efficient methods known for solving nonlinear least squares problems

$$\min \frac{1}{2} F(x)^T F(x), \quad (1)$$

where  $F : \Omega \rightarrow \mathbb{R}^m$  is differentiable function,  $\Omega \subset \mathbb{R}^n$  is an open set and  $m \geq n$ . Formally, the Gauss-Newton's method is described as follows: Given a initial point  $x_0 \in \Omega$ , define

$$x_{k+1} = x_k - [F'(x_k)^T F'(x_k)]^{-1} F'(x_k)^T F(x_k), \quad k = 0, 1, \dots$$

---

\*IME/UFG, Campus II- Caixa Postal 131, CEP 74001-970 - Goiânia, GO, Brazil (E-mail:orizon@mat.ufg.br). The author was supported in part by CNPq Grant 473756/2009-9, CNPq Grant 302024/2008-5, PRONEX-Optimization(FAPERJ/CNPq) and FUNAPE/UFG.

<sup>†</sup>COPPE-Sistemas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21945-970, BR (E-mail:maxlng@cos.ufrj.br). The author was supported in part by CNPq Grant 473756/2009-9.

<sup>‡</sup>COPPE-Sistemas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21945-970, BR (Email: poliveir@cos.ufrj.br). This author was supported in part by CNPq.

The convergence of this method may fail or it even fail to be well defined. To ensure that the method is well defined and converges to a solution of (1), some conditions must be impose. For instance, the classical convergence analysis (see [1, 2]) requires that  $F'$  satisfies the Lipschitz condition and the initial iterate to be "close enough" the solution, but it cannot make us clearly see how big is the convergence radius of the ball.

In the last years, there are many papers dealing with the convergence of the Newton's methods, including the Gauss-Newton's method, by relaxing the assumption of Lipschitz continuity of the derivative (see [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]). Those works in addition to improving the convergence theory (this allows us estimate the convergence radius and to enlarge the range of application) also permit us unify two results.

Our aim in this paper is present a new local convergence analysis for Gauss-Newton's method under majorant condition introduced by Kantorovich [15], and used with successful by Ferreira and Gonçalves [5], Ferreira [6] and Ferreira and Svaiter [7] for studying Newton's method. In our analysis, the classical Lipschitz condition is relaxed using a majorant function. It is worth pointing out that this condition is equivalent to Wang's condition introduced in [9] and used by Chen, Li [3, 4] and Li, et al. [8] for studying Gauss-Newton and Newton's method. The convergence analysis presented provides a clear relationship between the majorant function, which relax the Lipschitz continuity of the derivate, and the function associated with the nonlinear least square problem, see for example Lemmas 13, 14 and 15. Thus, the results presented here has the conditions and proof of convergence simpler and more didactic. Also, as in Chen, Li [3], it allow us to obtain the biggest range for the uniqueness of solution and the optimal convergence radius for the method with respect to majorant function. Moreover, two unrelated previous results pertaining Gauss-Newton's method are unified.

The organization of the paper is as follows. In Sect. 1.1, we list some notations and basic results used in our presentation. In Sect. 2 the main result is stated, and in Sect. 2.1 some properties involving the majorant function are established. In Sect. 2.2 we presented the relationships between the majorant function and the non-linear function  $F$ , and in Sect. 2.3 the optimal ball of convergence and the uniqueness of solution of convergence are established. In Sect. 2.4 the main result is proved and some applications of this result are given in Sect. 3.

## 1.1 Notation and auxiliary results

The following notations and results are used throughout our presentation. The open and closed ball at  $a \in \mathbb{R}^n$  and radius  $\delta > 0$  are denoted, respectively by

$$B(a, \delta) = \{x \in \mathbb{R}^n; \|x - a\| < \delta\}, \quad B[a, \delta] = \{x \in \mathbb{R}^n; \|x - a\| \leq \delta\}.$$

Let  $\mathbb{R}^{m \times n}$  denote the set of all  $m \times n$  matrix  $A$ ,  $A^\dagger$  denote the Moore-Penrose inverse of matrix  $A$ , and if  $A$  has full rank (namely:  $\text{rank}(A) = \min(m, n) = n$ ) then  $A^\dagger = (A^T A)^{-1} A^T$ .

**Lemma 1.** (*Banach's Lemma*) *Let  $B \in \mathbb{R}^{m \times m}$  and  $I \in \mathbb{R}^{m \times m}$ , the identity operator. If  $\|B - I\| < 1$ , then  $B$  is invertible and  $\|B^{-1}\| \leq 1/(1 - \|B - I\|)$ .*

*Proof.* See the proof of Lemma 1, p. 189 of Smale [16] with  $A = I$  and  $c = \|B - I\|$ .  $\square$

**Lemma 2.** Suppose that  $A, E \in \mathbb{R}^{m \times n} (m \geq n)$ ,  $B = A + E$ ,  $\|EA^\dagger\| < 1$ ,  $\text{rank}(A)=n$ , then  $\text{rank}(B)=n$ .

*Proof.* In fact,  $B = A + E = (I + EA^\dagger)A$ , from the condition  $\|EA^\dagger\| < 1$ , we have of Lemma 1 that  $I + EA^\dagger$  is invertible. So  $\text{rank}(B)=\text{rank}(A)=n$ .  $\square$

**Lemma 3.** Suppose that  $A, E \in \mathbb{R}^{m \times n}$ ,  $B = A + E$ ,  $\|A^\dagger\|\|E\| < 1$ ,  $\text{rank}(A) = \text{rank}(B)$ , then

$$\|B^\dagger\| \leq \frac{\|A^\dagger\|}{1 - \|A^\dagger\|\|E\|}.$$

Moreover, if  $\text{rank}(A) = \text{rank}(B) = \min(m, n)$ , there holds

$$\|B^\dagger - A^\dagger\| \leq \frac{\sqrt{2}\|A^\dagger\|^2\|E\|}{1 - \|A^\dagger\|\|E\|}.$$

*Proof.* See Lema 5.1. on pp. 40 of Stewart [17] and Wedin [18].  $\square$

**Proposition 4.** If  $0 \leq t < 1$ , then  $\sum_{i=0}^{\infty} (i+2)(i+1)t^i = 2/(1-t)^3$ .

*Proof.* Take  $k = 2$  in Lemma 3, pp. 161 of Blum, et al. [19].  $\square$

Also, the following auxiliary results of elementary convex analysis will be needed:

**Proposition 5.** Let  $R > 0$ . If  $\varphi : [0, R) \rightarrow \mathbb{R}$  is convex, then

$$D^+\varphi(0) = \lim_{u \rightarrow 0+} \frac{\varphi(u) - \varphi(0)}{u} = \inf_{0 < u} \frac{\varphi(u) - \varphi(0)}{u}.$$

*Proof.* See Theorem 4.1.1 on pp. 21 of Hiriart-Urruty and Lemaréchal [20].  $\square$

**Proposition 6.** Let  $\epsilon > 0$  and  $\tau \in [0, 1]$ . If  $\varphi : [0, \epsilon) \rightarrow \mathbb{R}$  is convex, then  $l : (0, \epsilon) \rightarrow \mathbb{R}$  define by

$$l(t) = \frac{\varphi(t) - \varphi(\tau t)}{t},$$

is increasing.

*Proof.* See Theorem 4.1.1 and Remark 4.1.2 on pp. 21 of Hiriart-Urruty and Lemaréchal [20].  $\square$

## 2 Local analysis for Gauss-Newton's method

Our goal is to state and prove a local theorem for Gauss-Newton's method. First, we will prove some results regarding the scalar majorant function, which relaxes the Lipschitz condition of the derivative of the function associated with the nonlinear least square problem. Then we will show that Gauss-Newton's method is well-defined and converges. We will also prove the uniqueness of the solution in a suitable region and the convergence rate will be established. The statement of the theorem is as follows:

**Theorem 7.** *Let  $\Omega \subseteq \mathbb{R}^n$  be an open set,  $F : \Omega \rightarrow \mathbb{R}^m$  a continuously differentiable function and  $m \geq n$ . Let  $x_* \in \Omega$ ,  $R > 0$  and*

$$c := \|F(x_*)\|, \quad \beta := \|[F'(x_*)^T F'(x_*)]^{-1} F'(x_*)^T\|, \quad \kappa := \sup \{t \in [0, R) : B(x_*, t) \subset \Omega\}.$$

*Suppose that  $x_*$  is a solution of (1),  $F'(x_*)$  has full rank and there exists a  $f : [0, R) \rightarrow \mathbb{R}$  continuously differentiable such that*

$$\|F'(x) - F'(x_* + \tau(x - x_*))\| \leq f'(\|x - x_*\|) - f'(\tau\|x - x_*\|), \quad (2)$$

*for all  $\tau \in [0, 1]$ ,  $x \in B(x_*, \kappa)$  and*

**h1)**  $f(0) = 0$  and  $f'(0) = -1$ ;

**h2)**  $f'$  is convex and strictly increasing;

**h3)**  $\sqrt{2}c\beta^2 D^+ f'(0) < 1$ .

*Let be given the positive constants  $\nu := \sup \{t \in [0, R) : \beta[f'(t) + 1] < 1\}$ ,*

$$\rho := \sup \left\{ t \in (0, \nu) : \frac{\beta[t f'(t) - f(t)] + \sqrt{2}c\beta^2[f'(t) + 1]}{t[1 - \beta(f'(t) + 1)]} < 1 \right\}, \quad r := \min \{\kappa, \rho\}.$$

*Then, the Gauss-Newton's method for solving (1), with starting point  $x_0 \in B(x_*, r) \setminus \{x_*\}$*

$$x_{k+1} = x_k - [F'(x_k)^T F'(x_k)]^{-1} F'(x_k)^T F(x_k), \quad k = 0, 1, \dots, \quad (3)$$

*is well defined, the generated sequence  $\{x_k\}$  is contained in  $B(x_*, r)$ , converges to  $x_*$  and*

$$\begin{aligned} \|x_{k+1} - x_*\| &\leq \frac{\beta[f'(\|x_0 - x_*\|)\|x_0 - x_*\| - f(\|x_0 - x_*\|)]}{\|x_0 - x_*\|^2[1 - \beta(f'(\|x_0 - x_*\|) + 1)]} \|x_k - x_*\|^2 \\ &\quad + \frac{\sqrt{2}c\beta^2[f'(\|x_0 - x_*\|) + 1]}{\|x_0 - x_*\|[1 - \beta(f'(\|x_0 - x_*\|) + 1)]} \|x_k - x_*\|, \quad k = 0, 1, \dots \end{aligned} \quad (4)$$

*Moreover, if  $[\beta(\rho f'(\rho) - f(\rho)) + \sqrt{2}c\beta^2(f'(\rho) + 1)]/[\rho(1 - \beta(f'(\rho) + 1))] = 1$  and  $\rho < \kappa$ , then  $r = \rho$  is the best possible convergence radius.*

*If, additionally,*

**h4)**  $2c\beta_0 D^+ f'(0) < 1$ , then the point  $x_*$  is the unique solution of (1) in  $B(x_*, \sigma)$ , where  
 $0 < \sigma := \sup\{t \in (0, \kappa) : [\beta(f(t)/t + 1) + c\beta_0(f'(t) + 1)/t] < 1\}$ ,  $\beta_0 := \|[F'(x_*)^T F'(x_*)]^{-1}\|$ .

**Remark 1.** The inequality (4) shows that if  $c = 0$  (the so-called zero-residual case), then the Gauss-Newton's method is locally  $Q$ -quadratically convergent to  $x_*$ . This behavior is quite similar to that of Newton's method (see [6, 9]). If  $c$  is small relative (the so-called small-residual case), the inequality (4) implies that the Gauss-Newton's method is locally  $Q$ -linearly convergent to  $x_*$ . However, if  $c$  is large (the so-called large-residual case), the Gauss-Newton's method may not be locally convergent at all, see condition **h3** and also example 10.2.4 on pp.225 of [1]. Hence, we may conclude that the Gauss-Newton's method perform better on zero-or small-residual problems than on large-residual problems, while the Newton's method is equally effective in all these cases.

For the zero-residual problems, i.e.,  $c = 0$ , the Theorem 7 becomes:

**Corollary 8.** Let  $\Omega \subseteq \mathbb{R}^n$  be an open set,  $F : \Omega \rightarrow \mathbb{R}^m$  a continuously differentiable function and  $m \geq n$ . Let  $x_* \in \Omega$ ,  $R > 0$  and

$$\beta := \|[F'(x_*)^T F'(x_*)]^{-1} F'(x_*)^T\|, \quad \kappa := \sup\{t \in [0, R) : B(x_*, t) \subset \Omega\}.$$

Suppose that  $F(x_*) = 0$ ,  $F'(x_*)$  has full rank and there exists a  $f : [0, R) \rightarrow \mathbb{R}$  continuously differentiable such that

$$\|F'(x) - F'(x_* + \tau(x - x_*))\| \leq f'(\|x - x_*\|) - f'(\tau\|x - x_*\|),$$

for all  $\tau \in [0, 1]$ ,  $x \in B(x_*, \kappa)$  and

**h1)**  $f(0) = 0$  and  $f'(0) = -1$ ;

**h2)**  $f'$  is convex and strictly increasing.

Let be given the positive constants  $\nu =: \sup\{t \in [0, \nu) : \beta[f'(t) + 1] < 1\}$ ,

$$\rho := \sup\{t \in (0, \nu) : [\beta(t f'(t) - f(t))]/[t(1 - \beta(f'(t) + 1))] < 1\}, \quad r := \min\{\kappa, \rho\}.$$

Then, the Gauss-Newton's method for solving (1), with initial point  $x_0 \in B(x_*, r) \setminus \{x_*\}$

$$x_{k+1} = x_k - [F'(x_k)^T F'(x_k)]^{-1} F'(x_k)^T F(x_k), \quad k = 0, 1, \dots,$$

is well defined, the sequence generated  $\{x_k\}$  is contained in  $B(x_*, r)$  and converges to  $x_*$  which is the unique solution of (1) in  $B(x_*, \sigma)$ , where  $0 < \sigma := \sup\{0 < t < \kappa : \beta[f(t)/t + 1] < 1\}$ . Moreover, there holds

$$\|x_{k+1} - x_*\| \leq \frac{\beta[f'(\|x_0 - x_*\|)\|x_0 - x_*\| - f(\|x_0 - x_*\|)]}{\|x_0 - x_*\|^2 [1 - \beta(f'(\|x_0 - x_*\|) + 1)]} \|x_k - x_*\|^2, \quad k = 0, 1, \dots$$

If, additionally,  $[\beta(\rho f'(\rho) - f(\rho))]/[\rho(1 - \beta(f'(\rho) + 1))] = 1$  and  $\rho < \kappa$ , then  $r = \rho$  is the best possible convergence radius.

**Remark 2.** When  $m = n$ , the Corollary 8 is similar to the result on Newton's method for solving nonlinear equations  $F(x) = 0$ , which has been obtained by Ferreira [6] in Theorem 2.1.

In order to prove Theorem 7 we need some results. From here on, we assume that all assumptions of Theorem 7 hold.

## 2.1 The majorant function

Our first goal is to show that the constant  $\kappa$  associated with  $\Omega$  and the constants  $\nu$ ,  $\rho$  and  $\sigma$  associated with the majorant function  $f$  are positive. Also, we will prove some results related to the function  $f$ .

We begin by noting that  $\kappa > 0$ , because  $\Omega$  is an open set and  $x_* \in \Omega$ .

**Proposition 9.** *The constant  $\nu$  is positive and there holds*

$$\beta[f'(t) + 1] < 1, \quad t \in (0, \nu).$$

*Proof.* As  $f'$  is continuous in  $(0, R)$  and  $f'(0) = -1$ , it is easy to conclude that

$$\lim_{t \rightarrow 0} \beta[f'(t) + 1] = 0.$$

Thus, there exists a  $\delta > 0$  such that  $\beta(f'(t) + 1) < 1$  for all  $t \in (0, \delta)$ . Hence,  $\nu > 0$ .

Using **h2** and definition of  $\nu$  the last part of the proposition follows. □

**Proposition 10.** *The following functions are increasing:*

- i)  $[0, R) \ni t \mapsto 1/[1 - \beta(f'(t) + 1)]$ ;
- ii)  $(0, R) \ni t \mapsto [tf'(t) - f(t)]/t^2$ ;
- iii)  $(0, R) \ni t \mapsto [f'(t) + 1]/t$ ;
- iv)  $(0, R) \ni t \mapsto f(t)/t$ .

As a consequence, are increasing the following functions

$$(0, R) \ni t \mapsto \frac{tf'(t) - f(t)}{t^2[1 - \beta(f'(t) + 1)]}, \quad (0, R) \ni t \mapsto \frac{f'(t) + 1}{t[1 - \beta(f'(t) + 1)]}.$$

*Proof.* The item **i** is immediate, because  $f'$  is strictly increasing in  $[0, R)$ .

For proving item **ii**, note that after some simple algebraic manipulations we have

$$\frac{tf'(t) - f(t)}{t^2} = \int_0^1 \frac{f'(t) - f'(\tau t)}{t} d\tau.$$

So, applying Proposition 6 with  $f' = \varphi$  and  $\epsilon = R$  the statement follows.

For establishing item **iii** use **h2**,  $f'(0) = -1$  and Proposition 6 with  $f' = \varphi$ ,  $\epsilon = R$  and  $\tau = 0$ .

Assumption **h2** implies that  $f$  is convex. As  $f(0) = 0$ , we have  $f(t)/t = [f(t) - f(0)]/[t - 0]$ . Hence, item **iv** follows by applying Proposition 6 with  $f = \varphi$  and  $\tau = 0$ .

For proving that the functions in the last part are increasing combine item **i** with **ii** for the first function and **i** with **iii** for the second function.  $\square$

**Proposition 11.** *The constant  $\rho$  is positive and there holds*

$$\frac{\beta[tf'(t) - f(t)] + \sqrt{2}c\beta^2[f'(t) + 1]}{t[1 - \beta(f'(t) + 1)]} < 1, \quad \forall t \in (0, \rho).$$

*Proof.* First, using **h1** and some algebraic manipulation gives

$$\frac{\beta[tf'(t) - f(t)] + \sqrt{2}c\beta^2[f'(t) + 1]}{t[1 - \beta(f'(t) + 1)]} = \frac{\beta \left[ f'(t) - \frac{f(t) - f(0)}{t - 0} \right] + \sqrt{2}c\beta^2 \frac{f'(t) - f'(0)}{t - 0}}{1 - \beta(f'(t) + 1)}.$$

Combing last equation with the assumption that  $f'$  is convex, we obtain from Proposition 5 that

$$\lim_{t \rightarrow 0} \frac{\beta[tf'(t) - f(t)] + \sqrt{2}c\beta^2[f'(t) + 1]}{t[1 - \beta(f'(t) + 1)]} = \sqrt{2}c\beta^2 D^+ f'(0).$$

Now, using **h3**, i.e.,  $\sqrt{2}c\beta^2 D^+ f'(0) < 1$ , we conclude that there exists a  $\delta > 0$  such that

$$\frac{\beta[tf'(t) - f(t)] + \sqrt{2}c\beta^2[f'(t) + 1]}{t[1 - \beta(f'(t) + 1)]} < 1, \quad t \in (0, \delta),$$

Hence,  $\delta \leq \rho$ , which prove the first statement.

For concluding the proof, we use the definition of  $\rho$ , above inequality and last part of Proposition 10.  $\square$

**Proposition 12.** *The constant  $\sigma$  is positives and there holds*

$$\beta(f(t)/t + 1) + c\beta_0(f'(t) + 1)/t < 1, \quad t \in (0, \sigma).$$

*Proof.* For proving that  $\sigma > 0$  we need the assumption **h4**. First, note that condition **h1** implies

$$\beta \left[ \frac{f(t)}{t} + 1 \right] + c\beta_0 \frac{f'(t) + 1}{t} = \beta \left[ \frac{f(t) - f(0)}{t - 0} - f'(0) \right] + c\beta_0 \frac{f'(t) - f'(0)}{t - 0}.$$

Therefore, using last equality together with the assumption that  $f'$  is convex and **h4** we have  $\lim_{t \rightarrow 0} [\beta(f(t)/t + 1) + c\beta_0(f'(t) + 1)/t] = c\beta_0 D^+ f'(0) < 1/2$ . Thus, there exists a  $\delta > 0$  such that

$$\beta \left[ \frac{f(t)}{t} + 1 \right] + c\beta_0 \frac{f'(t) + 1}{t} < 1, \quad t \in (0, \delta).$$

Hence,  $\delta \leq \sigma$ , which prove the first statement.

For concluding the proof, we use the definition of  $\sigma$ , above inequality and items **iii** and **iv** in Proposition 10.  $\square$

## 2.2 Relationship of the majorant function with the non-linear function

In this section we will present the main relationships between the majorant function  $f$  and the function  $F$  associated with the nonlinear least square problem.

**Lemma 13.** *Let  $x \in \Omega$ . If  $\|x - x_*\| < \min\{\nu, \kappa\}$ , then  $F'(x)^T F'(x)$  is invertible and the following inequalities hold*

$$\|[F'(x)^T F'(x)]^{-1} F'(x)^T\| \leq \frac{\beta}{1 - \beta[f'(\|x - x_*\|) + 1]},$$

and

$$\|[F'(x)^T F'(x)]^{-1} F'(x)^T - [F'(x_*)^T F'(x_*)]^{-1} F'(x_*)^T\| < \frac{\sqrt{2}\beta^2[f'(\|x - x_*\|) + 1]}{1 - \beta[f'(\|x - x_*\|) + 1]}.$$

In particular,  $F'(x)^T F'(x)$  is invertible in  $B(x_*, r)$ .

*Proof.* Let  $x \in \Omega$  such that  $\|x - x_*\| < \min\{\nu, \kappa\}$ . Since  $\|x - x_*\| < \nu$ , using the definition of  $\beta$ , the inequality (2) and last part of Proposition 9 we have

$$\|F'(x) - F'(x_*)\| \|[F'(x_*)^T F'(x_*)]^{-1} F'(x_*)^T\| \leq \beta[f'(\|x - x_*\|) - f'(0)] < 1.$$

For simply the notations define the following matrices

$$A = F'(x_*), \quad B = F'(x), \quad E = F'(x) - F'(x_*). \quad (5)$$

The last definitions together with latter inequality imply that

$$\|EA^\dagger\| \leq \|E\| \|A^\dagger\| < 1,$$

which, using that  $F'(x_*)$  has full rank, implies in view of Lemma 2 that  $F'(x)$  has full rank. So,  $F'(x)^T F'(x)$  is invertible and by definition of  $r$  we obtain that  $F'(x)^T F'(x)$  is invertible for all  $x \in B(x_*, r)$ .

We already knows that  $\text{rank} F'(x) = \text{rank} F'(x_*) = n$ . Hence, for concluding the lemma, first use definitions in (5) to obtain that  $\text{rank}(B) = \text{rank}(A) = n$  and then combine the above inequality and Lemma 3.  $\square$

Now, it is convenient to study the linearization error of  $F$  at point in  $\Omega$ , for that we define

$$E_F(x, y) := F(y) - [F(x) + F'(x)(y - x)], \quad y, x \in \Omega. \quad (6)$$

We will bound this error by the error in the linearization on the majorant function  $f$

$$e_f(t, u) := f(u) - [f(t) + f'(t)(u - t)], \quad t, u \in [0, R]. \quad (7)$$

**Lemma 14.** *If  $\|x - x_*\| < \kappa$ , then there holds  $\|E_F(x, x_*)\| \leq e_f(\|x - x_*\|, 0)$ .*

*Proof.* Since  $B(x_*, \kappa)$  is convex, we obtain that  $x_* + \tau(x - x_*) \in B(x_*, \kappa)$ , for  $0 \leq \tau \leq 1$ . Thus, as  $F$  is continuously differentiable in  $\Omega$ , definition of  $E_F$  and some simple manipulations yield

$$\|E_F(x, x_*)\| \leq \int_0^1 \| [F'(x) - F'(x_* + \tau(x - x_*))] \| \|x_* - x\| d\tau.$$

From the last inequality and the assumption (2), we obtain

$$\|E_F(x, x_*)\| \leq \int_0^1 [f'(\|x - x_*\|) - f'(\tau\|x - x_*\|)] \|x - x_*\| d\tau.$$

Evaluating the above integral and using definition of  $e_f$ , the statement follows.  $\square$

Lemma 13 guarantees, in particular, that  $F'(x)^T F'(x)$  is invertible in  $B(x_*, r)$  and consequently, the Gauss-Newton iteration map is well-defined. Let us call  $G_F$ , the Gauss-Newton iteration map for  $F$  in that region:

$$\begin{aligned} G_F : B(x_*, r) &\rightarrow \mathbb{R}^n \\ x &\mapsto x - [F'(x)^T F'(x)]^{-1} F'(x)^T F(x). \end{aligned} \quad (8)$$

One can apply a *single* Gauss-Newton iteration on any  $x \in B(x_*, r)$  to obtain  $G_F(x)$  which may not belong to  $B(x_*, r)$ , or even may not belong to the domain of  $F$ . So, this is enough to guarantee well definedness of only one iteration. To ensure that Gauss-Newton iterations may be repeated indefinitely, we need following result.

**Lemma 15.** *Let  $x \in \Omega$ . If  $\|x - x_*\| < r$ , then  $G_F$  is well defined and there holds*

$$\begin{aligned} \|G_F(x) - x_*\| &\leq \frac{\beta[f'(\|x - x_*\|)\|x - x_*\| - f(\|x - x_*\|)]}{\|x - x_*\|^2[1 - \beta(f'(\|x - x_*\|) + 1)]} \|x - x_*\|^2 \\ &\quad + \frac{\sqrt{2}c\beta^2[f'(\|x - x_*\|) + 1]}{\|x - x_*\|[1 - \beta(f'(\|x - x_*\|) + 1)]} \|x - x_*\|. \end{aligned}$$

*In particular,*

$$\|G_F(x) - x_*\| < \|x - x_*\|.$$

*Proof.* First note that, as  $\|x - x_*\| < r$  it follows from Lemma 13 that  $F'(x)^T F'(x)$  is invertible, then  $G_F(x)$  is well defined. Since  $F'(x_*)^T F(x_*) = 0$ , some algebraic manipulation and (8) yield

$$\begin{aligned} G_F(x) - x_* &= [F'(x)^T F'(x)]^{-1} F'(x)^T [F'(x)(x - x_*) - F(x) + F(x_*)] \\ &\quad + [F'(x_*)^T F'(x_*)]^{-1} F'(x_*)^T F(x_*) - [F'(x)^T F'(x)]^{-1} F'(x)^T F(x_*). \end{aligned}$$

From the last equation, properties of the norm and (6), we obtain

$$\begin{aligned} \|G_F(x) - x_*\| &\leq \left\| [F'(x)^T F'(x)]^{-1} F'(x)^T \right\| \|E_F(x, x_*)\| \\ &\quad + \left\| [F'(x_*)^T F'(x_*)]^{-1} F'(x_*)^T - [F'(x)^T F'(x)]^{-1} F'(x)^T \right\| \|F(x_*)\|. \end{aligned}$$

Since  $c = \|F(x_*)\|$ , combining last inequality with Lemmas 13 and 14 we have

$$\|G_F(x) - x_*\| \leq \frac{\beta e_f(\|x - x_*\|, 0)}{1 - \beta(f'(\|x - x_*\|) + 1)} + \frac{\sqrt{2}c\beta^2(f'(\|x - x_*\|) + 1)}{1 - \beta(f'(\|x - x_*\|) + 1)}.$$

Now, using (7) and **h1**, we conclude from last inequality that

$$\|G_F(x) - x_*\| \leq \frac{\beta[f'(\|x - x_*\|)\|x - x_*\| - f(\|x - x_*\|)]}{1 - \beta(f'(\|x - x_*\|) + 1)} + \frac{\sqrt{2}c\beta^2[f'(\|x - x_*\|) + 1]}{1 - \beta(f'(\|x - x_*\|) + 1)},$$

which is equivalent to the first inequality of the lemma.

To end the proof first note that the right hand side of the first inequality of the lemma is equivalent to

$$\left[ \frac{\beta[f'(\|x - x_*\|)\|x - x_*\| - f(\|x - x_*\|)]}{\|x - x_*\|[1 - \beta(f'(\|x - x_*\|) + 1)]} + \frac{\sqrt{2}c\beta^2[f'(\|x - x_*\|) + 1]}{\|x - x_*\|[1 - \beta(f'(\|x - x_*\|) + 1)]} \right] \|x - x_*\|.$$

On the other hand, as  $x \in B(x_*, r)/\{x_*\}$ , i.e.,  $0 < \|x - x_*\| < r \leq \rho$  we apply the Proposition 11 with  $t = \|x - x_*\|$  to conclude that the quantity in the bracket above is less than one. So, the last inequality of the lemma follows.  $\square$

## 2.3 Optimal ball of convergence and uniqueness

In this section, we will obtain the optimal convergence radius and the uniqueness of the solution.

**Lemma 16.** *If  $(\beta(\rho f'(\rho) - f(\rho)) + \sqrt{2}c\beta^2(f'(\rho) + 1))/\rho(1 - \beta(f'(\rho) + 1)) = 1$  and  $\rho < \kappa$ , then  $r = \rho$  is the best possible.*

*Proof.* Define the function  $h : (-\kappa, \kappa) \rightarrow \mathbb{R}$  by

$$h(t) = \begin{cases} -t/\beta + t - f(-t), & t \in (-\kappa, 0], \\ -t/\beta + t + f(t), & t \in [0, \kappa). \end{cases} \quad (9)$$

It is straightforward to show that  $h(0) = 0$ ,  $h'(0) = -1/\beta$ ,  $h'(t) = -1/\beta + 1 + f'(|t|)$  and that

$$|h'(t) - h'(\tau t)| \leq f'(|t|) - f'(\tau|t|), \quad \tau \in [0, 1], \quad t \in (-\kappa, \kappa).$$

So,  $F = h$  satisfy all assumption of Theorem 7 with  $c = |h(0)| = 0$ . Thus, as  $\rho < \kappa$ , it suffices to show that the Gauss-Newton's method applied for solving (1), with  $F = h$  and starting point  $x_0 = \rho$  does not converges. Since  $c = 0$  our assumption becomes

$$(\beta(\rho f'(\rho) - f(\rho))/\rho(1 - \beta(f'(\rho) + 1))) = 1. \quad (10)$$

Hence the definition of  $h$  in (9) together with last equality yields

$$x_1 = \rho - \frac{h'(\rho)^T h(\rho)}{h'(\rho)^T h'(\rho)} = \rho - \frac{-\rho/\beta + \rho + f(\rho)}{-1/\beta + 1 + f'(\rho)} = -\rho \left( \frac{\beta(\rho f'(\rho) - f(\rho))}{\rho(1 - \beta(f'(\rho) + 1))} \right) = -\rho.$$

Again, definition of  $h$  in (9) and assumption (10) gives

$$x_2 = -\rho - \frac{h'(-\rho)^T h(-\rho)}{h'(-\rho)^T h'(-\rho)} = -\rho - \frac{\rho/\beta - \rho - f(\rho)}{-1/\beta + 1 + f'(\rho)} = \rho \left( \frac{\beta(\rho f'(\rho) - f(\rho))}{\rho(1 - \beta(f'(\rho) + 1))} \right) = \rho.$$

Therefore, Gauss-Newton's method, for solving (1) with  $F = h$  and staring point  $x_0 = \rho$ , produces the cycle

$$x_0 = \rho, \quad x_1 = -\rho, \quad x_2 = \rho, \quad \dots,$$

as a consequence, it does not converge. Therefore, the lemma is proved.  $\square$

**Lemma 17.** *If additionally, **h4** holds, then the point  $x_*$  is the unique solution of (1) in  $B(x_*, \sigma)$ .*

*Proof.* Assume that  $y \in B(x_*, \sigma)$ ,  $y \neq x_*$  is also a solution of (1). Since  $F'(y)^T F(y) = 0$ , we have

$$y - x_* = y - x_* - [F'(x_*)^T F'(x_*)]^{-1} F'(y)^T F(y).$$

Using  $F'(x_*)^T F(x_*) = 0$ , after some algebraic manipulation the above equality becomes

$$\begin{aligned} y - x_* &= [F'(x_*)^T F'(x_*)]^{-1} F'(x_*)^T [F'(x_*)(y - x_*) - F(y) + F(x_*)] \\ &\quad + [F'(x_*)^T F'(x_*)]^{-1} (F'(x_*)^T - F'(y)^T) F(y). \end{aligned}$$

Combining the last equation with properties of the norm and definitions of  $c$ ,  $\beta$  and  $\beta_0$ , we obtain

$$\|y - x_*\| \leq \beta \int_0^1 \|F'(x_*) - F'(x_* + u(y - x_*))\| \|y - x_*\| du + c\beta_0 \|F'(x_*)^T - F'(y)^T\|.$$

Using (2) with  $x = x_* + u(y - x_*)$  and  $\tau = 0$  in the first term of the right-hand side, and  $x = y$  and  $\tau = 0$  in the second term of the right-hand side in last inequality, we have

$$\|y - x_*\| \leq \beta \int_0^1 [f'(u\|y - x_*\|) - f'(0)] \|y - x_*\| du + c\beta_0 [f'(\|y - x_*\|) - f'(0)].$$

Evaluating the above integral and using **h1**, the latter inequality becomes

$$\|y - x_*\| \leq \left( \beta \left[ \frac{f(\|y - x_*\|)}{\|y - x_*\|} + 1 \right] + c\beta_0 \left[ \frac{f'(\|y - x_*\|) + 1}{\|y - x_*\|} \right] \right) \|y - x_*\|,$$

Since  $0 < \|y - x_*\| < \sigma$ , using Proposition 12 with  $t = \|y - x_*\|$ , we have  $\|y - x_*\| < \|y - x_*\|$ , which is a contradiction. Therefore,  $y = x_*$ .  $\square$

**Remark 3.** Note that in the above lemma we have used the fact that condition (2) holds only for  $\tau = 0$ .

## 2.4 Proof of Theorem 7

First of all, note that the equation in (3) together (8) imply that the sequence  $\{x_k\}$  satisfies

$$x_{k+1} = G_F(x_k), \quad k = 0, 1, \dots \quad (11)$$

*Proof.* Since  $x_0 \in B(x_*, r) \setminus \{x_*\}$ , i.e.,  $0 < \|x_0 - x_*\| < r$ , by combination of Lemma 13, last inequality in Lemma 15 and induction argument it is easy to see that  $\{x_k\}$  is well defined and remains in  $B(x_*, r)$ .

Now, our goal is to show that  $\{x_k\}$  converges to  $x_*$ . As,  $\{x_k\}$  is well defined and contained in  $B(x_*, r)$ , combining (11) with Lemma 15 we have

$$\begin{aligned} \|x_{k+1} - x_*\| &\leq \frac{\beta[f'(\|x_k - x_*\|)\|x_k - x_*\| - f(\|x_k - x_*\|)]}{\|x_k - x_*\|^2[1 - \beta(f'(\|x_k - x_*\|) + 1)]} \|x_k - x_*\|^2 \\ &\quad + \frac{\sqrt{2}c\beta^2[f'(\|x_k - x_*\|) + 1]}{\|x_k - x_*\|[1 - \beta(f'(\|x_k - x_*\|) + 1)]} \|x_k - x_*\|, \end{aligned}$$

for all  $k = 0, 1, \dots$ . Using again (11) and the second part of and Lemma 15 it easy to conclude that

$$\|x_k - x_*\| < \|x_0 - x_*\|, \quad k = 1, 2, \dots \quad (12)$$

Hence combining two last inequalities with last part of Proposition 10 we obtain that

$$\begin{aligned} \|x_{k+1} - x_*\| &\leq \frac{\beta[f'(\|x_0 - x_*\|)\|x_0 - x_*\| - f(\|x_0 - x_*\|)]}{\|x_0 - x_*\|^2[1 - \beta(f'(\|x_0 - x_*\|) + 1)]} \|x_k - x_*\|^2 \\ &\quad + \frac{\sqrt{2}c\beta^2[f'(\|x_0 - x_*\|) + 1]}{\|x_0 - x_*\|[1 - \beta(f'(\|x_0 - x_*\|) + 1)]} \|x_k - x_*\|, \end{aligned}$$

for all  $k = 0, 1, \dots$ , which is the inequality (4). Now, using (12) and last inequality we have

$$\|x_{k+1} - x_*\| \leq \left[ \frac{\beta[f'(\|x_0 - x_*\|)\|x_0 - x_*\| - f(\|x_0 - x_*\|)] + \sqrt{2}c\beta^2[f'(\|x_0 - x_*\|) + 1]}{\|x_0 - x_*\|[1 - \beta(f'(\|x_0 - x_*\|) + 1)]} \right] \|x_k - x_*\|,$$

for all  $k = 0, 1, \dots$ . Applying Proposition 11 with  $t = \|x_0 - x_*\|$  it is straightforward to conclude from latter inequality that  $\{\|x_k - x_*\|\}$  converges to zero. So,  $\{x_k\}$  converges to  $x_*$ . The optimal convergence radius was proved in Lemma 16 and the last statement of theorem was proved in Lemma 17.  $\square$

### 3 Special cases

In this section, we present two special cases of Theorem 7. They include the classical convergence theorem on Gauss-Newton's method under Lipschitz condition and Smale's theorem on Gauss-Newton for analytical functions.

#### 3.1 Convergence result for Lipschitz condition

In this section we show a correspondent theorem to Theorem 7 under Lipschitz condition (see [1] and [2] ) instead of the general assumption (2).

**Theorem 18.** *Let  $\Omega \subseteq \mathbb{R}^n$  be an open set,  $F : \Omega \rightarrow \mathbb{R}^m$  be continuously differentiable in  $\Omega$  and  $m \geq n$ . Let  $x_* \in \Omega$  and*

$$c := \|F(x_*)\|, \quad \beta := \|[F'(x_*)^T F'(x_*)]^{-1} F'(x_*)^T\|, \quad \kappa := \sup \{t \in [0, R) : B(x_*, t) \subset \Omega\}.$$

*Suppose that  $x_*$  is a solution of (1),  $F'(x_*)$  has full rank and there exists a  $K > 0$  such that*

$$\sqrt{2}c\beta^2 K < 1, \quad \|F'(x) - F'(y)\| \leq K\|x - y\|, \quad \forall x, y \in B(x_*, \kappa).$$

*Let*

$$r := \min \left\{ \kappa, (2 - 2\sqrt{2}K\beta^2 c)/(3K\beta) \right\}.$$

*Then, the Gauss-Newton methods for solving (1), with initial point  $x_0 \in B(x_*, r)/\{x_*\}$*

$$x_{k+1} = x_k - [F'(x_k)^T F'(x_k)]^{-1} F'(x_k)^T F(x_k), \quad k = 0, 1, \dots$$

*is well defined, the sequence generated  $\{x_k\}$  is contained in  $B(x_*, r)$ , converges to  $x_*$  and*

$$\|x_{k+1} - x_*\| \leq \frac{\beta K}{2(1 - \beta K\|x_0 - x_*\|)} \|x_k - x_*\|^2 + \frac{\sqrt{2}c\beta^2 K}{1 - \beta K\|x_0 - x_*\|} \|x_k - x_*\|, \quad k = 0, 1, \dots$$

*Moreover, if  $(2 - 2\sqrt{2}K\beta^2 c)/(3K\beta) < \kappa$ , then  $r = (2 - 2\sqrt{2}K\beta^2 c)/(3K\beta)$  is the best possible convergence radius.*

*If, additionally,  $2c\beta_0 K < 1$ , then the point  $x_*$  is the unique solution of (1) in  $B(x_*, (2 - 2c\beta_0 K)/(\beta K))$ , where  $\beta_0 := \|[F'(x_*)^T F'(x_*)]^{-1}\|$ .*

*Proof.* It is immediate to prove that  $F$ ,  $x_*$  and  $f : [0, \kappa) \rightarrow \mathbb{R}$  defined by  $f(t) = Kt^2/2 - t$ , satisfy the inequality (2), conditions **h1** and **h2**. Since  $\sqrt{2}c\beta^2K < 1$  and  $2c\beta_0K < 1$  the conditions **h3** and **h4** also hold. In this case, it is easy to see that the constants  $\nu$  and  $\rho$  as defined in Theorem 7, satisfy

$$0 < \rho = (2 - 2\sqrt{2}K\beta^2c)/(3K\beta) \leq \nu = 1/\beta K,$$

as a consequence,  $0 < r = \min\{\kappa, \rho\}$ . Moreover, it is straightforward to show that

$$[\beta(\rho f'(\rho) - f(\rho)) + \sqrt{2}c\beta^2(f'(\rho) + 1)]/[\rho(1 - \beta(f'(\rho) + 1))] = 1,$$

and  $[\beta(f(t)/t + 1) + c\beta_0(f'(t) + 1)/t] < 1$  for all  $t \in (0, (2 - 2c\beta_0K)/(\beta K))$ . Therefore, as  $F$ ,  $r$ ,  $f$  and  $x_*$  satisfy all of the hypotheses of Theorem 7, taking  $x_0 \in B(x_*, r) \setminus \{x_*\}$  the statements of the theorem follow from Theorem 7.  $\square$

For the zero-residual problems, i.e.,  $c = 0$ , the Theorem 18 becomes:

**Corollary 19.** *Let  $\Omega \subseteq \mathbb{R}^n$  be an open set,  $F : \Omega \rightarrow \mathbb{R}^m$  be continuously differentiable in  $\Omega$  and  $m \geq n$ . Let  $x_* \in \Omega$  and*

$$\beta := \|[F'(x_*)^T F'(x_*)]^{-1} F'(x_*)^T\|, \quad \kappa := \sup \{t \in [0, R) : B(x_*, t) \subset \Omega\}.$$

*Suppose that  $F(x_*) = 0$ ,  $F'(x_*)$  has full rank and there exists a  $K > 0$  such that*

$$\|F'(x) - F'(y)\| \leq K\|x - y\|, \quad \forall x, y \in B(x_*, \kappa).$$

*Let*

$$r := \min\{\kappa, 2/(3K\beta)\}.$$

*Then, the Gauss-Newton methods for solving (1), with initial point  $x_0 \in B(x_*, r) \setminus \{x_*\}$*

$$x_{k+1} = x_k - [F'(x_k)^T F'(x_k)]^{-1} F'(x_k)^T F(x_k), \quad k = 0, 1, \dots,$$

*is well defined, the sequence generated  $\{x_k\}$  is contained in  $B(x_*, r)$  and converges to  $x_*$  which is the unique solution of (1) in  $B(x_*, 2/(\beta K))$ . Moreover, there holds*

$$\|x_{k+1} - x_*\| \leq \frac{\beta K}{2(1 - \beta K\|x_0 - x_*\|)} \|x_k - x_*\|^2, \quad k = 0, 1, \dots$$

*If, additionally  $2/(3K\beta) < \kappa$ , then  $r = 2/(3K\beta)$  is the best possible convergence radius.*

**Remark 4.** *When  $m = n$ , the Corollary 19 merge in the results on the Newton's method for solving nonlinear equations  $F(x) = 0$ , which has been obtained by Ferreira [6] in Theorem 3.1 and Remark 3.3.*

### 3.2 Convergence result under Smale's condition

In this section we present a correspondent theorem to Theorem 7 under Smale's condition. For more details about Smale's condition see [16].

**Theorem 20.** *Let  $\Omega \subseteq \mathbb{R}^n$  be an open set,  $F : \Omega \rightarrow \mathbb{R}^m$  an analytic function and  $m \geq n$ . Let  $x_* \in \Omega$  and*

$$c := \|F(x_*)\|, \quad \beta := \|[F'(x_*)^T F'(x_*)]^{-1} F'(x_*)^T\|, \quad \kappa := \sup\{t > 0 : B(x_*, t) \subset \Omega\}.$$

*Suppose that  $x_*$  is a solution of (1),  $F'(x_*)$  has full rank and*

$$\gamma := \sup_{n \geq 1} \left\| \frac{F^{(n)}(x_*)}{n!} \right\|^{1/(n-1)} < +\infty, \quad 2\sqrt{2}c\beta^2\gamma < 1. \quad (13)$$

*Let  $a := (2 + 3\beta - \sqrt{2}c\beta^2\gamma)$ ,  $b := 4(1 + \beta)(1 - 2\sqrt{2}c\beta^2\gamma)$  and*

$$r := \min \left\{ \kappa, (a - \sqrt{a^2 - b}) / (2\gamma(1 + \beta)) \right\}.$$

*Then, the Gauss-Newton methods for solving (1), with initial point  $x_0 \in B(x_*, r) \setminus \{x_*\}$*

$$x_{k+1} = x_k - [F'(x_k)^T F'(x_k)]^{-1} F'(x_k)^T F(x_k), \quad k = 0, 1, \dots,$$

*is well defined, the sequence generated  $\{x_k\}$  is contained in  $B(x_*, r)$ , converges to  $x_*$  and*

$$\begin{aligned} \|x_{k+1} - x_*\| &\leq \frac{\beta\gamma}{(1 - \gamma\|x_0 - x_*\|)^2 - \beta\gamma(2\|x_0 - x_*\| - \gamma\|x_0 - x_*\|^2)} \|x_k - x_*\|^2 \\ &\quad + \frac{\sqrt{2}c\beta^2\gamma(2 - \gamma\|x_0 - x_*\|)}{(1 - \gamma\|x_0 - x_*\|)^2 - \beta\gamma(2\|x_0 - x_*\| - \gamma\|x_0 - x_*\|^2)} \|x_k - x_*\|, \quad k = 0, 1, \dots \end{aligned}$$

*Moreover, if  $(a - \sqrt{a^2 - b}) / (2\gamma(1 + \beta)) < \kappa$ , then  $r = (a - \sqrt{a^2 - b}) / (2\gamma(1 + \beta))$  is the best possible convergence radius.*

*If additionally,  $4c\beta_0\gamma < 1$ , then the point  $x_*$  is the unique solution (1) in  $B(x_*, \sigma)$ , where  $\sigma := (\omega_1 - \sqrt{\omega_1^2 - \omega_2}) / (2\gamma(1 + \beta))$ ,  $\omega_1 := (2 + \beta - c\beta_0)$ ,  $\omega_2 := 4(1 + \beta)(1 - 2c\beta_0\gamma)$ ,  $\beta_0 := \|[F'(x_*)^T F'(x_*)]\|$ .*

We need the following result to prove the above theorem.

**Lemma 21.** *Let  $\Omega \subseteq \mathbb{R}^n$  be an open set and  $F : \Omega \rightarrow \mathbb{R}^m$  an analytic function. Suppose that  $x_* \in \Omega$  and  $B(x_*, 1/\gamma) \subset \Omega$ , where  $\gamma$  is defined in (13). Then, for all  $x \in B(x_*, 1/\gamma)$  there holds*

$$\|F''(x)\| \leq (2\gamma) / (1 - \gamma\|x - x_*\|)^3.$$

*Proof.* Let  $x \in \Omega$ . Since  $F$  is an analytic function, we have

$$F''(x) = \sum_{n=0}^{\infty} \frac{1}{n!} F^{(n+2)}(x_*) (x - x_*)^n.$$

Combining (13) and the above equation we obtain, after some simple calculus, that

$$\|F''(x)\| \leq \gamma \sum_{n=0}^{\infty} (n+2)(n+1)(\gamma\|x - x_*\|)^n.$$

On the other hand, as  $B(x_*, 1/\gamma) \subset \Omega$  we have  $\gamma\|x - x_*\| < 1$ . So, from Proposition 4 we conclude

$$\frac{2}{(1 - \gamma\|x - x_*\|)^3} = \sum_{n=0}^{\infty} (n+2)(n+1)(\gamma\|x - x_*\|)^n.$$

Combining the two above equations, we obtain the desired result.  $\square$

The next result gives a condition that is easier to check than condition (2), whenever the functions under consideration are twice continuously differentiable.

**Lemma 22.** *Let  $\Omega \subseteq \mathbb{R}^n$  be an open set,  $x_* \in \Omega$  and  $F : \Omega \rightarrow \mathbb{R}^m$  be twice continuously on  $\Omega$ . If there exists a  $f : [0, R) \rightarrow \mathbb{R}$  twice continuously differentiable such that*

$$\|F''(x)\| \leq f''(\|x - x_*\|), \quad (14)$$

*for all  $x \in \Omega$  such that  $\|x - x_*\| < R$ . Then  $F$  and  $f$  satisfy (2).*

*Proof.* Taking  $\tau \in [0, 1]$  and  $x \in \Omega$ , such that  $x_* + \tau(x - x_*) \in \Omega$  and  $\|x - x_*\| < R$ , we obtain that

$$\| [F'(x) - F'(x_* + \tau(x - x_*))] \| \leq \int_{\tau}^1 \|F''(x_* + t(x - x_*))\| \|x - x_*\| dt.$$

Now, as  $\|x - x_*\| < R$  and  $f$  satisfies (14), we obtain from the last inequality that

$$\| [F'(x) - F'(x_* + \tau(x - x_*))] \| \leq \int_{\tau}^1 f''(t\|x - x_*\|) \|x - x_*\| dt.$$

Evaluating the latter integral, the statement follows.  $\square$

**[Proof of Theorem 20].** Consider the real function  $f : [0, 1/\gamma) \rightarrow \mathbb{R}$  defined by

$$f(t) = \frac{t}{1 - \gamma t} - 2t.$$

It is straightforward to show that  $f$  is analytic and that

$$f(0) = 0, \quad f'(t) = 1/(1 - \gamma t)^2 - 2, \quad f'(0) = -1, \quad f''(t) = (2\gamma)/(1 - \gamma t)^3, \quad f^n(0) = n! \gamma^{n-1},$$

for  $n \geq 2$ . It follows from the last equalities that  $f$  satisfies **h1** and **h2**. Since  $2\sqrt{2}c\beta^2\gamma < 1$  and  $4c\beta_0\gamma < 1$  the conditions **h3** and **h4** also hold. Now, as  $f''(t) = (2\gamma)/(1 - \gamma t)^3$  combining Lemmas 22, 21 we conclude that  $F$  and  $f$  satisfy (2) with  $R = 1/\gamma$ . In this case, it is easy to see that the constants  $\nu$  and  $\rho$  as defined in Theorem 7, satisfy

$$0 < \rho = (a - \sqrt{a^2 - b})/(2\gamma(1 + \beta)) < \nu = ((1 + \beta) - \sqrt{\beta(1 + \beta)})/(\gamma(1 + \beta)) < 1\gamma,$$

and as a consequence,  $0 < r = \min\{\kappa, \rho\}$ . Moreover, it is not hard to see that

$$[\beta(\rho f'(\rho) - f(\rho)) + \sqrt{2}c\beta^2(f'(\rho) + 1)]/[\rho(1 - \beta(f'(\rho) + 1))] = 1,$$

and  $[\beta(f(t)/t + 1) + c\beta_0(f'(t) + 1)/t] < 1$  for all  $t \in (0, \sigma)$ . Therefore, as  $F$ ,  $\sigma$ ,  $f$  and  $x_*$  satisfy all hypothesis of Theorem 7, taking  $x_0 \in B(x_*, r) \setminus \{x_*\}$ , the statements of the theorem follow from Theorem 7.  $\square$

For the zero-residual problems, i.e.,  $c = 0$ , the Theorem 20 becomes:

**Corollary 23.** *Let  $\Omega \subseteq \mathbb{R}^n$  be an open set,  $F : \Omega \rightarrow \mathbb{R}^m$  an analytic function and  $m \geq n$ . Let  $x_* \in \Omega$ , and*

$$\beta := \|[F'(x_*)^T F'(x_*)]^{-1} F'(x_*)^T\|, \quad \kappa := \sup\{t > 0 : B(x_*, t) \subset \Omega\}.$$

*Suppose that  $F(x_*) = 0$ ,  $F'(x_*)$  has full rank and*

$$\gamma := \sup_{n \geq 1} \left\| \frac{F^{(n)}(x_*)}{n!} \right\|^{1/(n-1)} < +\infty.$$

*Let*

$$r := \min \left\{ \kappa, (2 + 3\beta - \sqrt{\beta(8 + 9\beta)})/(2\gamma(1 + \beta)) \right\}.$$

*Then, the Gauss-Newton methods for solving (1), with initial point  $x_0 \in B(x_*, r) \setminus \{x_*\}$*

$$x_{k+1} = x_k - [F'(x_k)^T F'(x_k)]^{-1} F'(x_k)^T F(x_k), \quad k = 0, 1, \dots,$$

*is well defined, is contained in  $B(x_*, r)$  and converges to  $x_*$  which is the unique solution of (1) in  $B(x_*, 1/(\gamma(1 + \beta)))$ . Moreover, there holds*

$$\|x_{k+1} - x_*\| \leq \frac{\beta\gamma}{(1 - \gamma\|x_0 - x_*\|)^2 - \beta\gamma(2\|x_0 - x_*\| - \gamma\|x_0 - x_*\|^2)} \|x_k - x_*\|^2, \quad k = 0, 1, \dots$$

*If, additionally,  $(2 + 3\beta - \sqrt{\beta(8 + 9\beta)})/(2\gamma(1 + \beta)) < \kappa$ , then  $r = ((2 + 3\beta - \sqrt{\beta(8 + 9\beta)})/(2\gamma(1 + \beta)))$  is the best possible convergence radius.*

**Remark 5.** *When  $m = n$ , the Corollary 23 is similar to the results on the Newton's method for solving nonlinear equations  $F(x) = 0$ , which has been obtained by Ferreira [6] in Theorem 3.4.*

## References

- [1] J. E. Dennis, Jr. and Robert B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. Corrected reprint of the 1983 original.
- [2] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer Series in Operations Research. Springer-Verlag, New York, 1999.
- [3] Jinhai Chen and Weiguo Li. Convergence of Gauss-Newton’s method and uniqueness of the solution. *Appl. Math. Comput.*, 170(1):686–705, 2005.
- [4] Jinhai Chen and Weiguo Li. Convergence behaviour of inexact Newton methods under weak Lipschitz condition. *J. Comput. Appl. Math.*, 191(1):143–164, 2006.
- [5] O.P. Ferreira and M. L. N. Goncalves. Local convergence analysis of inexact newton-like methods under majorant condition. *Comput. Optim. Appl.*, Article in Press:1–21, 2009.
- [6] Orizon P. Ferreira. Local convergence of Newton’s method in Banach space from the viewpoint of the majorant principle. *IMA J. Numer. Anal.*, 29(3):746–759, 2009.
- [7] O. P. Ferreira and B. F. Svaiter. Kantorovich’s majorants principle for Newton’s method. *Comput. Optim. Appl.*, 42(2):213–229, 2009.
- [8] Chong Li, Wen-Hong Zhang, and Xiao-Qing Jin. Convergence and uniqueness properties of Gauss-Newton’s method. *Comput. Math. Appl.*, 47(6-7):1057–1067, 2004.
- [9] Xinghua Wang. Convergence of Newton’s method and uniqueness of the solution of equations in Banach space. *IMA J. Numer. Anal.*, 20(1):123–134, 2000.
- [10] T. J. Ypma. Local convergence of inexact Newton methods. *SIAM J. Numer. Anal.*, 21(3):583–590, 1984.
- [11] Jinhai Chen. The convergence analysis of inexact gauss—newton methods for nonlinear problems. *Comput. Optim. Appl.*, 40(1):97–118, 2008.
- [12] Petko D. Proinov. General local convergence theory for a class of iterative processes and its applications to Newton’s method. *J. Complexity*, 25(1):38–62, 2009.
- [13] Petko D. Proinov. New general convergence theory for iterative processes and its applications to newton-kantorovich type theorems. *Journal of Complexity*, 26(1):3 – 42, 2010.

- [14] Chong Li, Nuchun Hu, and Jinhua Wang. Convergence behavior of gauss-newton's method and extensions of the smale point estimate theory. *Journal of Complexity*, In Press, Corrected Proof:–, 2010.
- [15] L. V. Kantorovič. The principle of the majorant and Newton's method. *Doklady Akad. Nauk SSSR (N.S.)*, 76:17–20, 1951.
- [16] Steve Smale. Newton's method estimates from data at one point. In *The merging of disciplines: new directions in pure, applied, and computational mathematics (Laramie, Wyo., 1985)*, pages 185–196. Springer, New York, 1986.
- [17] G. W. Stewart. On the continuity of the generalized inverse. *SIAM J. Appl. Math.*, 17:33–45, 1969.
- [18] Per-Åke Wedin. Perturbation theory for pseudo-inverses. *Nordisk Tidskr. Informationsbehandling (BIT)*, 13:217–232, 1973.
- [19] Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale. *Complexity and real computation*. Springer-Verlag, New York, 1998. With a foreword by Richard M. Karp.
- [20] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms. I*, volume 305 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. Fundamentals.